AMENDMENTS TO THE CLAIMS

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

1. (Currently amended) A system that facilitates speech recognition by modeling speech dynamics, comprising:

an input component that receives acoustic data; and

a model component that employs the acoustic data to characterize speech, the model component comprising model parameters that form a mapping relationship from unobserved speech dynamics to observed speech acoustics, the model parameters are employed to decode an unobserved phone sequence of speech based, at least in part, upon a variational learning technique;

wherein the model component is based, at least in part, upon a hidden dynamic model in the form of a segmental switching state space model, the segmental switching state space model comprises respective states having respective durations in time corresponding to soft boundaries of respective phones in the unobserved phone sequence.

2. (Original) The system of claim 1, modification of at least one of the model parameters being based upon a variational expectation maximization algorithm having an E-step and M-step.

3. (Original) The system of claim 2, modification of at least one of the model parameters being based, at least in part, upon a mixture of Gaussian (MOG) posteriors based on a variational technique.

4. (Currently amended) The system of claim 3, the model component being based, at least in part, upon:

$$q(s_{1:N}, x_{1:N}) = \prod_n q(x_n \mid s_n) q(s_n),$$

where x is a state of the model,

s is a phone index,

n is a frame number,

N is [[the]] a number of frames to be analyzed, and

q is a probability approximation.

5. (Original) The system of claim 2, modification of at least one of the model parameters being based, at least in part, upon a mixture of hidden Markov model (HMM) posteriors based on a variational technique.

6. (Previously presented) The system of claim 1, the model component selecting an approximate posterior distribution relating to the acoustic data and optimizing a posterior distribution by minimizing a Kullback-Leibler (KB) distance thereof to an exact posterior distribution.

7-8. (Canceled)

9. (Currently amended) The system of claim [[7]] 1, the model component employing, at least in part, the state equation:

$$x_n = A_s x_{n-1} + (I - A_s) u_s + w,$$

and the observation equation:

$$y_n = C_s x_n + c_s + v,$$

where n is a frame number,

s is a phone index,

x is the hidden dynamics,

y is an acoustic feature vector,

v is Gaussian white noise,

w is Gaussian white noise,

A is a phone dependent system matrix,

I is an identity matrix,

u is a target vector, and

C and c are [[the]] parameters for mapping from x to y.

10. (Currently amended) The system of claim [[7]] 1, the model component being expressed, at least in part, in terms of probability distributions:

$$p(s_n = s \mid s_{n-1} = s') = \pi_{s's},$$

$$p(\mathbf{x}_n \mid s_n = s, \mathbf{x}_{n-1}) = N(\mathbf{x}_n \mid \mathrm{A}_s \mathbf{x}_{n-1} + \mathrm{a}_s, \mathrm{B}_s),$$

$$p(\mathbf{y}_n \mid s_n = s, \mathbf{x}_n) = N(\mathbf{y}_n \mid \mathrm{C}_s \mathbf{x}_n + \mathrm{c}_s, \mathrm{D}_s),$$

where $\pi_{s's}$ is a phone transition probability matrix, $\mathbf{a}_s = (\mathbf{I} - \mathbf{A}_x)\mathbf{u}_s$, where A_x is a phone dependent system matrix, I is an identity matrix, and u is a target vector,

N denotes a Gaussian distribution with mean and precision matrix as the parameters,

A and a are [[the]] parameters for mapping from a state of x at a given frame to a state of x at an immediately following frame,

B represents [[the]] a covariance matrix of [[the]] a residual vector after the mapping from a state of x at a given frame to a state of x at an immediately following frame,

C and c are [[the]] parameters for mapping from x to y, and,

D represents [[the]] a covariance matrix of [[the]] a residual vector after the mapping from x to y.

11. (Canceled)

12. (Currently amended) A method that facilitates modeling speech dynamics in a speech recognition system comprising:

decoding an unobserved phone sequence of speech from acoustic data based, at least in part, upon a speech model, the speech model based upon a hidden dynamic model in the form of a segmental switching state space model, comprising one or more states corresponding to respective phones in the unobserved phone sequence having respective durations corresponding to estimated soft boundaries for the phones, and further comprising at least two sets of parameters, a first set of model parameters describing unobserved speech dynamics and a second set of model parameters describing a

relationship between an unobserved speech dynamic vector and an observed acoustic feature vector;

calculating a posterior distribution based on at least the first set of model parameters and the second set of model parameters; and,

modifying at least one of the model parameters based, at least in part, upon the calculated posterior distribution.

13. (Canceled)

14. (Currently amended) A method ~~that facilitates~~ of modeling speech dynamics from acoustic data for speech recognition comprising:

recovering a phone sequence of speech from acoustic data based, at least in part, upon a speech model, wherein the speech model is a segmental switching state space model and comprises a plurality of model parameters and one or more states corresponding to respective phones in the phone sequence created by segmenting the speech model in time based on estimated soft boundaries for the phones;

calculating an approximation of a posterior distribution based on the model parameters, the model parameters and the approximation based upon a mixture of Gaussians; and,

modifying at least one model parameter based, at least in part, upon the calculated approximated posterior distribution and minimization of a Kullback-Leibler distance of the approximation from an exact posterior distribution.

15. (Canceled)

16. (Currently amended) The method of claim 14, calculation of the approximation of the posterior distribution being based, at least in part, upon:

$$q(s_{1:N}, x_{1:N}) = \prod_n q(x_n \mid s_n) q(s_n),$$

where x is a state of the model,

s is a phone index,

n is a frame number,

N is [[the]] a number of frames to be analyzed, and

q is a posterior probability approximation.

17. (Currently amended) A method that facilitates ~~modeling~~ creating a model of speech dynamics for a speech recognition application comprising:

recovering a phone sequence of speech from acoustic data based, at least in part, upon a speech model in the form of a segmental switching state space model comprising one or more states respectively corresponding to the phone sequence, the states are generated by segmenting the speech model in time based on soft boundaries for respective phones in the phone sequence;

calculating an approximation of a posterior distribution based on model parameters, the model parameters and the approximation based upon a hidden Markov model posterior; and,

modifying at least one of the model parameters based, at least in part, upon the calculated approximated posterior distribution and minimization of a Kullback-Leibler distance of the approximation from an exact posterior distribution.

18. (Currently amended) The method of claim 17, calculation of the approximation of the posterior distribution being based, at least in part, upon:

$$q(s_{1:N}, x_{1:N}) = \prod_{n=1}^{N} q(x_n \mid s_n) \cdot \prod_{n=2}^{N} q(s_n \mid s_{n-1}) \cdot q(s_1)$$

where x is a state of the model,

s is a phone index,

n is a frame number,

N is [[the]] a number of frames to be analyzed, and

q is a posterior probability approximation.

19. (Currently amended) A data packet transmitted between two or more computer components that facilitates modeling of speech dynamics in a speech recognition application, the signal comprising:

a data structure associated with one or more recovered speech parameters; and

a segmental switching state space speech model that employs acoustic data and the one or more recovered speech parameters to facilitate modeling of speech dynamics and to recover a phone sequence of speech based on the reversed speech parameters, the phone sequence of speech including one or more phones respectively including ~~recovered speech parameters including~~ at least one articulation parameter and at least one duration parameter.

20. (Currently amended) A computer readable medium containing computer executable instructions operable to perform a method of modeling speech dynamics comprising:

receiving acoustic data;

modeling speech based on a segmental switching state space model comprising a first set of parameters that describe unobserved speech dynamics, [[and]] a second set of parameters that describe a relationship between the unobserved speech dynamic vector and an observed acoustic feature vector, and[[,]] a set of states having respective durations corresponding to soft phone boundaries determined from the acoustic data; and

modifying at least one of the first set of parameters and the second set of parameters based, at least in part, upon a variational learning technique.

21. (Currently amended) A system that facilitates modeling speech dynamics comprising:

means for receiving acoustic data; and,

means for characterizing speech as a segmental switching state space model based, at least in part, upon the acoustic data,

wherein the means for modeling speech employs model parameters that are modified based, at least in part, upon a variational learning technique and one or more states having respective durations corresponding to estimated soft phone boundaries.

22. (New) The system of claim 1, wherein the hidden dynamic model comprises a series of time-varying transition matrices based on the unobserved phone sequence to

constrain the durations of the respective states to the estimated soft boundaries of the respective phones in the unobserved phone sequence, thereby forcing the respective states to be consistent in time with the unobserved phone sequence.

23. (New) The system of claim 1, wherein the unobserved speech dynamics are vocal tract resonances associated with movement of an articulator.

24. (New) The system of claim 2, wherein the modification of at least one of the model parameters is based on a multimodal posterior distribution and a variational technique for processing the multimodal posterior distribution.

25. (New) The method of claim 12, wherein the unobserved speech dynamics comprise vocal tract resonance frequency parameters.

26. (New) The method of claim 12, wherein the calculating a posterior distribution includes calculating a multimodal posterior distribution based on the first set of model parameters and the second set of model parameters and the modifying includes modifying at least one of the model parameters based on the multimodal posterior distribution and calculus of variation.

27. (New) A method of modeling speech dynamics for a speech processing application, comprising:

constructing a speech model, the speech model is based on a hidden dynamic model in the form of a segmental switching state space model for speech applications, the constructing a speech model comprising:

initializing a first set of model parameters that describes unobserved vocal tract resonance frequencies;

initializing a second set of model parameters that describes a mapping relationship between the unobserved vocal tract resonance frequencies and observed acoustic data;

creating a state equation based on the first set of model parameters to express the unobserved vocal tract resonance frequencies as a set of states respectively corresponding to phones in an unobserved phonetic transcript, the state equation is a linear dynamic equation that describes transitions between states in the set of states in terms of a phone-dependent system matrix and a target vector and includes a first Gaussian noise parameter;

creating an observation equation that utilizes the first set of model parameters and the second set of model parameters to represent a phone-dependent mapping between the unobserved vocal tract resonance frequencies and the observed acoustic data, the mapping selected from the group consisting of a linear mapping and a piecewise linear mapping within respective phones, the observation equation includes a second Gaussian noise parameter;

estimating soft phone boundaries for phones in the unobserved phonetic transcript under an expectation-maximization (EM) framework; and

constructing a series of time-varying transition matrices based on the phonetic transcript to constrain the set of states to respective time durations corresponding to the estimated soft phone boundaries for phones in the phonetic transcript, thereby forcing the states to be consistent in time with the phonetic transcript;

calculating an estimated multimodal posterior distribution based on the constructed speech model, the first set of model parameters, and the second set of model parameters; and

modifying one or more model parameters to minimize a Kullback-Leibler distance from the estimated multimodal posterior distribution to an exact posterior distribution, the modifying is based on an EM framework having an expectation step of model inference and a maximization step of model learning, the model learning is based on a variational learning technique that employs calculus of variation.